

1 **CROSS REFERENCE TO RELATED APPLICATIONS**

2 This application claims the benefit of U.S. Provisional Application Ser. No.
3 60/269,872, which was filed with the U.S. Patent and Trademark Office on Feb. 16, 2001.

4 **BACKGROUND OF THE INVENTION**

5 1. Field of the Invention

6 The present invention relates to the isolation of a nucleic acid sequences that
7 encode an enzyme that catalyzes the transfer of electrons to protons for the production of
8 molecular hydrogen, and more particularly to iron hydrogenase and genes encoding for
9 the iron hydrogenase in microscopic organisms known as unicellular green algae.

11 2. Prior Art

12 Molecular hydrogen is currently being considered as a candidate for replacing or
13 supplementing fossil fuels and as a source of clean energy. A potential method for
14 producing hydrogen on a commercial scale is the photobiological production of hydrogen
15 by eukaryotic organisms. Green algae respond to anaerobic stress by switching the
16 oxidative pathway to a fermentative metabolism. The fermentation of organic compounds
17 and residual photosynthetic electron transport in the green algae are associated with
18 hydrogen evolution. The key enzyme hydrogenase, which is synthesized only after an
19 anaerobic adaptation, catalyzes the reversible reduction of protons to molecular hydrogen.
20 This method is capable of generating renewable hydrogen fuel from light and water,
21 which are among nature's most plentiful resources.

1 The ability of green algae, such as *Chlamydomonas reinhardtii*, to produce
2 hydrogen from water has been recognized for over 55 years. This reaction is catalyzed by
3 a reversible hydrogenase, an enzyme that is induced in the cells after exposure to a short
4 period of anaerobiosis. However, the activity of the hydrogenase is rapidly lost when
5 cells are illuminated because of the immediate inactivation of the reversible hydrogenase
6 by photosynthetically generated O_2 .

7 Methods have been devised to circumvent the hydrogenase inactivation problem.
8 US Pat. No. 4,532,210 discloses the biological production of hydrogen in an algal culture
9 using an alternating light and dark cycle. The process comprises alternating a step for
10 cultivating the alga in water under aerobic conditions in the presence of light to
11 accumulate photosynthetic products (starch) in the alga, and a step for cultivating the alga
12 in water under microaerobic conditions in the dark to decompose the accumulated
13 material by photosynthesis to evolve hydrogen. This method uses a nitrogen gas purge
14 technique to remove oxygen from the culture.

15 US Pat. No. 4,442,211 discloses that the efficiency of a process for producing
16 hydrogen, by subjecting algae in an aqueous phase to light irradiation, is increased by
17 culturing algae which has been bleached during a first period of irradiation in a culture
18 medium in an aerobic atmosphere until it has regained color and then subjecting this
19 algae to a second period of irradiation wherein hydrogen is produced at an enhanced rate.
20 A reaction cell is used wherein light irradiates the culture in an environment which is
21 substantially free of CO_2 and atmospheric O_2 . This environment is maintained by passing

1 an inert gas (e.g. helium) through the cell to remove all hydrogen and oxygen generated
2 by the splitting of water molecules in the aqueous medium. Although continuous purging
3 of H₂-producing cultures with inert gases has allowed for the sustained production of H₂,
4 such purging is expensive and impractical for large-scale mass cultures of algae. In view
5 of the foregoing, there remains a need for a microorganism that produces a hydrogenase
6 enzyme suitable for use in a sustainable process of photosynthetic hydrogen production.

8 SUMMARY

9 Accordingly, it is an object of the present invention to provide a gene encoding
10 for hydrogenase and a method for using the gene product for the microbial production of
11 molecular hydrogen. Specifically, the invention provides isolated nucleic acid sequences
12 encoding a stable hydrogenase enzyme (HydA) that will catalyze the reduction of protons
13 to form molecular hydrogen.

14 Another object of the present invention is to provide isolated nucleic acid
15 sequences encoding a protein that catalyzes the reduction of protons to form molecular
16 hydrogen comprising SEQ. ID. NO. 1. SEQ. ID. NO. 1 comprises a nucleic acid sequence
17 that encodes *Scenedesmus obliquus* HydA.

18 It is yet a further object of the present invention to provide isolated nucleic acid
19 sequences encoding a protein that catalyzes the reduction of protons to form molecular
20 hydrogen comprising SEQ. ID. NO. 2. SEQ. ID. NO. 2 comprises a nucleic acid sequence
21 that encodes *Chlamydomonas reinhardtii* HydA.

1 A further object of the present invention is to provide fragments of the nucleic
2 acid sequence comprising SEQ. ID. NO.1 or SEQ. ID. NO.2, encoding iron hydrogenase,
3 that code for products that maintain the biological activity necessary to catalyze the
4 transfer of electrons to protons in a process for producing molecular hydrogen. Such
5 fragments can be either recombinant or synthetic or a combination thereof.

6 The features of the invention believed to be novel are set forth with particularity
7 in the appended claims. However the invention itself, both as to organization and method
8 of operation, together with further objects and advantages thereof may be best be
9 understood by reference to the following description taken in conjunction with the
10 accompanying drawings in which:

11 BRIEF DESCRIPTION OF THE DRAWINGS

12 Figure 1 is a schematic representation of *S. obliquus* HydA genomic and cDNA
13 structures.

14 Figure 2 is a comparison of the *S. obliquus*-derived iron hydrogenase amino acid
15 sequence with HydA sequences derived from other organisms.

16 Figure 3 is a schematic diagram showing the conserved cysteine residues and
17 other important amino acids of the H cluster

18 Figure 4 (A) is a schematic view of the structure of *S. obliquus* HydA.

19 Figure 4 (B) is a schematic view of the *S. obliquus* electron donor ferredoxin.

20 Figure 5 (A) is a schematic map of the cDNA and the genomic DNA region of
21 HydA from *C. reinhardtii*.showing the structural features of the HydA cDNA. Coding

1 regions are marked as large arrows with the transit peptide shown in black. lines indicate
2 5' and 3' URTs.

3 Figure 5 (B) is a schematic map of the cDNA and the genomic DNA region of
4 HydA from *C. reinhardtii*. showing the structural features of the HydA cDNA. In Fig. 3
5 (B), the mosaic structure of hydA is illustrated by gray (exons) and white (introns) boxes.
6 The RNA and DNA probes that were used for the blotting experiments are noted on the
7 Figure.

8 Figure 6 shown the nucleotide sequence of the hydA cDNA and the deduced
9 amino acid sequence of the hydrogenase from *C. reinhardtii*.

10 Figure 7 is a schematic diagram showing the light-dependent photoevolution of
11 hydrogen in green algae.

12 DESCRIPTION OF THE PREFERRED EMBODIMENTS

13 The isolation, purification and biochemical and genetic characterization of a novel
14 iron hydrogenase from *S. obliquus* and *C. reinhardtii* and *C. fuscus* is disclosed.

15 I. *Scenedesmus obliquus*

16 *S. obliquus* Algal Strains and Growth Conditions

17 Wild-type *S. obliquus* Kützing 276-6 was obtained originally from the culture
18 collection of algae at the University of Göttingen. Cells were cultured
19 photoheterotrophically in batch cultures at 25 °C under continuous irradiance of 150
20 $\mu\text{mol photons per square meter per second}$. For anaerobic adaptation, 4-liter cultures were
21 bubbled with air supplemented with 5% CO_2 . After harvesting the cells in the mid-

1 exponential stage of growth, the pellet was resuspended in fresh Tris acetate phosphate
2 (TAP) medium. The algae were anaerobically adapted by flushing the culture with argon
3 in the dark.

4 **Hydrogen Evolution Assay**

5 The *in vitro* hydrogenase activity was measured by using a Hewlett Packard (HP
6 5890, Series II) gas chromatograph, equipped with a thermal conductivity detector and a
7 molecular sieve column. Methyl viologen reduced by sodium dithionite was used as an
8 electron donor. 1 unit is defined as the amount of hydrogenase evolving 1 μmol of
9 molecular hydrogen (H_2) per minute at 25 °C.

10 The *in vivo* activity in the presence of different inhibitors of the photosynthetic
11 electron flow was determined as described by Happe et al., in: *European Journal of*
12 *Biochemistry*, **214**, 475-481 (1993). After anaerobic adaptation, algal cells were
13 harvested, diluted in fresh Tris acetate phosphate medium, and transferred to sealed tubes.
14 Inhibitors were added 1 hour before H_2 evolving activity was measured. Cells were
15 broken by sonification. Thylakoid membranes and photosynthetic transport chain
16 remained intact as demonstrated by oxygen polarography. Ferredoxin of both *C.*
17 *reinhardtii* and *S. obliquus* was isolated according to the method of Schmitter et al. (*Eur.*
18 *J. of Biochem.*, **172**, pages 405-412 (1988)).

19 **Rapid Amplification of cDNA Ends-Polymerase Chain Reaction (RACE-PCR)**

20 RACE-PCR was performed with the Clontech SMARTTMRACE cDNA
21 Amplification Kit (Clontech Laboratories, Palo Alto, CA) according to the

1 manufacturer's recommendations, except for modification of the PCR and hybridization
2 conditions. Starting material consisted of 1 µg of mRNA from anaerobically adapted
3 cells. The reverse transcription reaction was carried out with a Moloney murine leukemia
4 virus reverse transcriptase in two separate reaction tubes containing either the 5' or the 3'
5 RACE-PCR specific primer from the kit. The cDNA of each sample served as template
6 for the following PCR. For the 5'-RACE-PCR, a Universal Primer Mix (UPM) and the
7 antisense primer, Sc7, were used. The amplification of the 3'-cDNA end was performed
8 with a UPM and the sense primer Sc6. To obtain more distinct PCR signals, the PCR
9 was repeated for both reactions with nested universal primers and designed primers
10 (inverse Sc6 and inverse Sc7, respectively) using a dilution of the products of the first
11 PCR as template.

12 **Primer Extension**

13 RACE-PCR was also implemented to map the transcription initiation site of the
14 *hydA* mRNA. A gene-specific primer (Sc17) was used to carry out the first strand cDNA
15 synthesis with the Superscript II reverse transcriptase (Life Technologies, Rockville, MD,
16 USA) and 200 ng of mRNA as template. PCR was performed using either Sc12 or Sc27
17 and the SMART™ specific adapter primer UPM. Two different DNA fragments of 234
18 bp and 183 bp were amplified under standard PCR conditions. Both fragments were
19 cloned into the pGEM®T-Easy vector (Promega, Madison, WI, USA) and sequenced
20 using primers from the polylinker of the vector.

21 **Genome Walking with Genomic DNA**

1 Applying the Clontech Walker Kit (Clontech Laboratories), genomic libraries
2 from *S. obliquus* were generated by digestion with different blunt-end cutting
3 endonucleases (*NaeI*, *DraI*, *PvuII*, *HincII* and *EcoRV*) and by adapter ligation at the
4 ends of the resulting DNA fragments. These libraries were utilized as independent
5 templates in five different PCR reactions. Two gene-specific primers (Sc27, Sc35)
6 derived from the *hydA* cDNA sequence of *S. Obliquus* were used in combination with a
7 kit adapter primer (AP1) in a first PCR reaction. Subsequently, 1 µl of the first PCR
8 served as a template in a secondary PCR, applying two nested gene-specific primers (i-
9 Sc10, Sc32) along with a nested kit adapter primer (AP2). The resulting products were
10 cloned into **pGEMTMT-Easy** and sequenced. Sequencing was performed by the dideoxy
11 termination method (see, for example: Sanger et al., *Proc. Natl. Acad. Sci., U.S.A.*, **74**,
12 5463-5467 (1977)).

13 **Purification of the Fe-hydrogenase**

14 40-liter cultures of *Scenedesmus obliquus* were grown heterotrophically. After
15 centrifugation (10 min, 5000 x g) the pellet we re-suspended in 200 ml TAP medium.
16 The cells were anaerobically adapted by flushing the solution with argon for 1 hour in the
17 dark. All further purification steps were performed in an anaerobic chamber (Coylab,
18 Ann Arbor, MI, USA). The cells were disrupted in a 50 mM Tris/HCl, buffer pH 8.0, 10
19 mM sodium dithionite by vortexing 3 min with glass beads. The further purification
20 steps were made as described hereinbelow for the isolation of the Fe-hydrogenase of
21 *Chlamydomonas reinhardtii*. Automated Edman degradation of the N-terminal site of the

1 protein was performed with an Applied Biosystem model 477A sequencer with online
2 analyser model 120 A.

3 **RNA Blot Hybridization**

4 Total RNA of *S. Obliquus* was isolated according to the method described by
5 Johanningmeier et al. (*J. Biol. Chem.* **259**, 13541-13549 (1984)). Equal amounts (20 µg)
6 were separated electrophoretically on 1.2% agarose gels containing formaldehyde. The
7 RNA was transferred onto nylon membranes (Hybond⁺, Amersham) and hybridized with
8 RNA probes labeled DIG-dUPT using *in vitro* transcription methodology. A 1.3 kb
9 *EcoRI* cDNA fragment was used to detect transcripts with a *hydA* gene, while a DIG-
10 dUPT- labeled cDNA encoding constitutively expressed plastocyanin, was used as a
11 control. Hybridization reactions were carried out using protocols supplied by the
12 manufacturer (Roche Diagnostics, Mannheim, Germany).

13 **Sequence Analysis Software**

14 Nucleic acid and protein sequences were analyzed with the programs Sci Ed
15 Central (Scientific Educational Software) and ClustalW. The Blast server of the National
16 Center for Biotechnology Information (Bethesda, MD) was used for database searches.

17 **Recombinant Expression in *E.coli***

18 The *hydA* open reading frame was amplified by PCR using the primer pair Sc29
19 and Ac30 containing flanking *NdeI*-*BamHI* sites. The PCR product was cloned into the
20 **pGEMTM**-Easy vector. After digestion with *NdeI*-*BamHI*, the *hydA* gene was cloned
21 into the corresponding site of the pET9a expression vector (Promega), producing

1 pLF29.2. The insert of pLF29.2 was sequenced confirming that the fragment contained
 2 the exact full coding region of the hydrogenase without transit peptide. *E.coli* strain
 3 BL21(DE3)pLysS was transformed with pLF29.2. Expression was induced with 1 mM
 4 isopropyl-thio- β -D-galactoside at an OD₆₀₀ of 0.3. Pelleted cells were re-suspended in
 5 lysis buffer (100 mM Tris/HCl; 4 mM EDTA; 16% Glycine; 2% SDS; 2%
 6 Mercaptoethanol; 0.05% Bromophenolblue; 8 M Urea). After heating, the protein extract
 7 was separated by 10% SDS-PAGE and blotted onto a PVDF membrane. Western blot
 8 analyses were performed using antisera against the Fe-hydrogenase of *Chlamydomonas*
 9 *reinhardtii* at 1:1000 dilution.

10 **Results for *Scenedesmus obliquus***

11 **Induction of Hydrogenase Activity and Purification of the Fe-hydrogenase Protein**

12 Anaerobic adaptation is the most efficient way to induce hydrogenase activity in
 13 *Scenedesmus obliquus*. Bubbling the alga culture in the dark with argon led to a dramatic
 14 increase (10-fold) of hydrogenase activity during the first 2 hours. The enzyme of *S.*
 15 *obliquus* was purified to homogeneity by successive column chromatography. Since the
 16 enzyme is irreversibly inactivated by very low levels of oxygen, all purification steps
 17 were performed under strictly anaerobic conditions and in the presence of reducing agents
 18 (dithionite). The purification scheme resulted in a 5200-fold purification of HydA with
 19 5% recovery (data not shown). The most powerful step for purifying the protein was a Q-
 20 Sepharose high performance column chromatography with pH gradient elution. Gel
 21 infiltration chromatography of hydrogenase on a calibrated Superdex-75 column resulted

1 in a single activity peak corresponding to a molecular mass of 45kDa. The monomeric
2 structure of the enzyme could also be shown on a SDS polyacrylamide gel after
3 Coomassie-blue staining (data not shown). The N-terminal sequence of HydA was
4 determined by Edman degradation. The protein sequence
5 (AGPTAECDRPPAPAPKAXHWQ) is, except for two amino acids, identical to the
6 amino acid sequence deduced from the DNA data (AGPTAECDCPPAPAPKAPHWQ).
7 In the course of the purification procedure there was no indication of a second
8 hydrogenase in *S. obliquus* because the hydrogenase activity was never separated into
9 distinct fractions. Biochemical data show a high similarity of HydA to the Fe-
10 hydrogenase from *C. reinhardtii* (Table 1). The enzymes have a high temperature
11 optimum of about 60 °C, are strongly inhibited by O₂ and CO, and catalyze the H₂-
12 evolution with a typical high specific activity. Experiments with inhibitors of translators
13 on ribosomes (data not shown) and analysis of the gene structure show that *HydA* from *S.*
14 *obliquus* is translated in the cytoplasm and transported to the chloroplast.

15 Table I

16
17 Biochemical data comparison of purified iron hydrogenases from
18 *C. reinhardtii* and *S. obliquus*

19	<i>C. reinhardtii</i>	<i>S. obliquus</i>
20 Size	49 kDa	44.5 kDa
21 Specific activity	935 U/mg protein	700 U/mg protein
22 Temperature optimum	60 °C	60 °C
23 pH optimum	6.9	7.3
24 Localization	chloroplast stroma	chloroplast
25 Coding site	nuclear	nuclear
26 pI value	5.3	5.17
27 <i>K_M</i> value (MV)	830 μM	800 μM
28 <i>K_M</i> value (ferredoxin)	35 μM	Not determined

Ferredoxin Is the Natural Electron Donor of the Fe-hydrogenase

Hydrogenase activity was determined in intact and broken cells after anaerobic adaptation. The integrity of the photosynthetic electron transport in the sonified cell preparation was demonstrated by the rate of oxygen evolution ($154 \mu\text{mole O}_2/\text{mg Chl} \times \text{h}$). This rate corresponds to 85% of the oxygen evolution measure with intact *Scenedesmus* cells.

In *S. obliquus*, the hydrogen evolution is linked to the photosynthetic electron transport chain through PSI. As shown in Table II, the cells were still able to photoproduce hydrogen when electron flow on the PSII was blocked by DCMU. In contrast, addition of DCMIB resulted in inhibition of the H_2 -production, thus giving evidence of the involvement of PSI in the supply of electrons to hydrogenase. With reference to Table II, after anaerobic adaptation, cells were harvested, diluted in fresh TAP medium, and incubated with inhibitors as described herein. α -PetF-antibody was raised against spinach ferredoxin. In Table II, DCMU = 3-(3,4-dichlorophenyl)-1,1-dimethylurea; DBMIB = 2,5-dibromo-3-methyl-6-isopropyl-p-benzochinone; Sulfo-DSPD = sulfodisalicylidinepropanediamin; DCPIP = 2,6-dichlorophenolindophenol. The electron transport from PSI to ferredoxin was inhibited using the artificial electron acceptor DCPIP. In this reaction, DCPIP is reduced instead of ferredoxin and electron transfer to hydrogenase is interrupted.

Table II

Effects of different photosynthetic inhibitors on hydrogenase activity

		Hydrogenase activity
		Units/mg chlorophyll
Intact cells (control)		0.11
+ DCMU (10^{-5} M)		0.10
+ DBMIB (10^{-5} M)		0.005
Broken cells		0.1
+ DCMU (10^{-5} M)		0.11
+ DBMIB (10^{-5} M)		0.006
+ DCPIP (10^{-4} M)		0.003
+ sulfo-DSPD (10^{-4} M)		0.003
+ α -PetF-antibody (1:1000)		0.008

Hydrogenase activity was dramatically reduced (up to 30-fold) by the ferredoxin antagonist sulfo-DSPD (Table II). Similar results were achieved with α -PetF-antibodies that specifically recognize the ferredoxin protein. In both cases, the hydrogenase enzyme cannot evolve hydrogen, thus demonstrating the role of ferredoxin as the obligatory donor for the hydrogenase reaction.

The electron transfer properties of different plant-type ferredoxins were measured *in vitro* with dithionite as a reducing reagent. The ferredoxin proteins of spinach, *C. reinhardtii* and *S. obliquus* were comparable regarding their capability to reduce purified *S. obliquus* hydrogenase. In this assay, H_2 -evolving activities of 420, 390 and 350 U/mg protein with *S. obliquus*, *C. reinhardtii* and spinach ferredoxin, respectively, were observed. No hydrogen production could be measured with other possible electron donors like cytochrome and NADPH. In *D. desulfuricans*, the Fe-hydrogenase has been

1 reported to catalyze both hydrogen production and uptake with low potential multiheme
2 cytochromes such as cytochrome c_3 .

3 **Molecular Characterization of *hdyA* Encoding a Fe-hydrogenase**

4 In order to isolate the gene encoding a Fe-hydrogenase in *S. obliquus*, polyA⁺
5 RNA was isolated from cell cultures after one hour of anaerobic adaptation. Isolated
6 RNA was transcribed and amplified by RT-PCR using oligonucleotides derived from
7 conserved regions within the *C. reinhardtii* *HydA* gene (Happe, unpublished results).
8 The complete cDNA clone of 2609 bp was obtained by 5'- and 3'- RACE PCR. It
9 contains an open reading frame of 1344 bp encoding a polypeptide of 448 amino acids
10 (Fig. 1) followed by an extensive 3' UTR of about 1100 bp. The coding region of *S.*
11 *obliquus* *hdyA* exhibits features common to other green algae such as high GC content
12 (64.2%) and a characteristic putative polyadenylation signal, TGTAAG, 15 bp upstream of
13 the polyA⁺ sequence.

14 In order to examine the exon-intron structure and the promoter region of the *hdyA*
15 gene, about 5 kb of the genomic DNA from *S. obliquus* were sequenced. The gene
16 comprises 5 introns with a total size of 1310 bp (Fig. 1) whose 5'- and 3'- end contain
17 typical plant splice donor and acceptor sites that follow the GT/AG rule. In Figure 1,
18 which is a schematic representation of *S. obliquus* *hdyA* genomic and cDNA structures,
19 the coding region of the *hdyA* cDNA is illustrated as a *large arrow* with the transit
20 peptide shown in *black*. The untranslated 59 and 39 sequences are marked as *lines*. The
21 *arrows* below indicate the sequencing strategy; each *arrow* represents an independent

1 sequence determination. *TSP*, transcription start point; *ATG*, start codon. The mosaic
2 structure of *hydA* is indicated by *gray* (exons) and *white* (introns) boxes. The S2 probe
3 and different restriction enzymes that were used in the Southern blot experiments are
4 indicated on the figure.

5 A genomic southern blot was probed with a 750 bp PCR fragment to determine
6 the copy number of the *hydA* gene. Single bands were observed in lanes with samples
7 digested with *HincII*, *EcoRV* and *NdeI* and a double band in the lane containing genomic
8 DNA digested with *SacI*. The band migration positions matched the sizes predicted from
9 the sequence of the *hydA* gene, indicating that HydA is encoded by a single copy gene.
10 The same hybridization pattern was observed even under low stringency conditions
11 (hybridization temperature 50 °C; data not shown). The transcription start position was
12 determined by primer extension using RACE-PCR and was found 139 bp upstream of the
13 ATG start codon. Several primers within 100 bp of the 5'-end of the known *hydA* cDNA
14 were designed to confirm the accuracy of the transcription initiation site. All of the
15 sequenced PCR clones had the same 5'-ends at position +1. As described for other green
16 algae genes, a highly conserved TATA box element upstream of the transcription start
17 point is absent (see, for example: C.D. Sifflow in: *The Molecular Biology of Chloroplasts*
18 *and Mitochondria in Chlamydomonas*, pp 25-40, (J.D. Goldschmidt-Clermont et al., eds.)
19 Kluwer Academic Publishers, Dordrecht, The Netherlands (1998)). However, the
20 TACATAT motive at position -25 in a GC rich region shows similarities to other TATA
21 motives in *C. reinhardtii* and therefore might be involved in gene expression.

HydA Is a Novel Type of Fe-hydrogenase

The polypeptide derived from the cDNA sequence has a length of 448 amino acids and a predicted molecular mass of 48.5 kDa (44.5 kDa excluding the transit peptide); consequently HydA is the smallest hydrogenase protein known so far. The N-terminus of HydA is basic and contains numerous hydroxylated amino acids and a Val-X-Ala motive at position 35, a characteristic feature of chloroplast transit peptides. The processed HydA protein is compared with four bacterial and two eukaryotic Fe-hydrogenases as shown in Figure 2. Figure 2 is a comparison of the *S. obliquus*-derived iron hydrogenase amino acid sequence with HydA sequences derived from other organisms. In Figure 2, the protein alignment was done by using the Vector NTI program (InforMax). *White letters with black background* indicate amino acids identical to the HydA protein. *Black letters with gray background* indicate conserved changes of the amino acids. *S. o.*, *S. obliquus* (this work); *M. e.*, *Megasphaera elsdenii*; *D. d.*, *D. desulfuricans*; *T. v.*, *Trichomonas vaginalis*; *C. p.*, *C. pasteurianum*; *T. m.*, *T. maritima*; *N. o.*, *N. ovalis*.

The homology in the carboxy-terminal region of all proteins is quite striking. For example, the *S. obliquus* HydA protein shows 44% identity and 57% similarity to the *C. pasteurianum* Fe-hydrogenase. The H-cluster in *S. obliquus* might be coordinated by four cysteine residues at positions 120, 175, 335, and 340. Other strictly conserved amino acid structures such as FTSCCPGW (334-350), TGGVMEAALR (474-483) and MACPGGCXXGGGQP (586-589) probably define a pocket surrounding the active

1 center as shown by the structural data of *C. pasteurianum* and *D. desulfuricans*. On the
2 other hand, the N-terminal region is completely different from all other Fe-hydrogenases.
3 The protein sequences of the other enzymes comprise at least two [4Fe-4S] ferredoxin-
4 like domains (called "f-cluster") which are necessary for the electron transport from the
5 electron donor to the catalytic center. The Fe-hydrogenases of *C. pasteurianum*,
6 *Thermotoga maritima* and *Nyctotherus ovalis* contain an extra [4Fe-4S] cluster and one
7 [2Fe-2S] center. This N-terminal domain with the F-cluster or other [Fe-S] centers is
8 completely lacking in HydA of *S. obliquus*. This indicates that there is a direct electron
9 transport pathway from the exogenous donor to the H-cluster.

10 To verify that the isolated cDNA encodes a Fe-hydrogenase, the *hydA* clone was
11 expressed in the heterologous system *E. coli*. One band of recombinant HydA was
12 observed on SDS-PAGE at approximately 44kDa, in agreement with the molecular mass
13 of the polypeptide predicted from the cDNA sequence. Antibodies raised against the
14 HydA protein of the *C. reinhardtii*, which cross-react with other Fe-hydrogenases but not
15 with NiFe-hydrogenases (data not shown), were applied in Western blot analysis. One
16 distinct signal with the over-expressed HydA protein of *S. obliquus* was obtained.
17 The lysate of induced *E. coli* cells exhibited no hydrogenase activity. This result
18 corresponds to observations by Voordouw *et al.* (*Eur. J. Biochem.*, **162**, 31-36 (1987))
19 and Stokkermans *et al.* (*FEMS Microbiol. Lett.*, **49**, 217-222 (1989)) who also detected
20 no H₂-production of recombinant Fe-hydrogenases in *E. coli* cells. The reason for that

1 might be that the bacterial cells do not have the ability to assemble the special H-cluster
2 of the Fe-hydrogenases.

3 **Rapid Induction of *hydA* mRNA during Anaerobic Adaptation**

4 The regulation of the *hydA* gene expression was examined by Northern blot
5 analysis and reverse transcription-PCR (RT-PCR). Aerobically grown cells of *S.*
6 *obliquus* did not show a hydrogenase activity. Total RNA and mRNA were isolated from
7 cells which were induced by argon bubbling for 0, 1 and 4 hours. Northern blot analysis
8 and RT-PCR demonstrated that the *hydA* gene is expressed after anaerobic adaptation.
9 There is a very weak signal without adaptation (t=0), but strong signals of the transcript
10 could be detected after anaerobic induction. The full length of the *hydA* cDNA clone was
11 confirmed by the transcript signal (2.6 kb) on the Northern blot.

12 **Discussion**

13 In green algae, the occurrence of a hydrogen metabolism induced by anaerobic
14 conditions is well established. Despite the great interest in hydrogen evolution for
15 practical applications ("biophotolysis"), the hydrogenase genes from green algae have
16 heretofore not been isolated. The *hydA* gene and the HydA protein of *Scenedemus*
17 *obliquus* presented herein belong to the class of Fe-hydrogenases.

18 Fe-hydrogenases have been isolated only from certain anaerobic bacteria and
19 some anaerobic eukaryotes as well as from the anaerobically adapted green alga *C.*
20 *reinhardtii* (T. Happe et al., *Eur. J. Biochem.* **214**, 475-481 (1993)). The enzymes are

1 found to exist in monomeric, dimeric and multimeric forms; however, in eukaryotes, only
2 monomeric proteins have been isolated.

3 The HydA protein of *S. obliquus* is synthesized in the cytoplasm. The first 35
4 residues M¹ to A³⁵) of the amino acid sequence derived from the cDNA sequence are
5 supposed to function as a short transit peptide which routes the nuclear encoded protein to
6 the chloroplast. Several positively charged amino acids which describe a typical feature
7 for algal transit peptides are found in HydA. The three terminal residues of the signal
8 sequence, Val-X-Ala, constitute the consensus for stromal peptidases.

9 The hydrogenase of *S. obliquus* represents a novel type of Fe-hydrogenase. The
10 monomeric enzyme of 448 amino acids and a calculated molecular mass of 44.5 kDA for
11 the processed protein is the smallest Fe-hydrogenase isolate so far. The protein sequence
12 consists of an unusual N-terminal domain and a large carboxyterminal domain containing
13 the catalytic site. The structurally important C-terminus of the *S. obliquus* HydA
14 sequence is very similar to that of other Fe-hydrogenases. Four cysteine residues at
15 positions C¹²⁰, C¹⁷⁵, C³³⁶ and C³⁴⁰ coordinate the special [6Fe] cluster (H-cluster) of the
16 active site. A number of addition residues define the environment of the catalytic center.
17 Peters *et al.* postulated twelve amino acids in *C. pasteurianum* to form a hydrophobic
18 pocket around the cofactor (Science, **282**, 1853-1858 (1998)). Ten residues are strictly
19 conserved while two amino acids vary within the Fe- hydrogenase family (S²³², I²⁶⁸, in *C.*
20 *pasteurianum*, A¹¹⁹, T¹⁵⁵ in *T. vaginalis* and A⁴⁴, T⁸⁰ in *S. obliquus*). A small insertion of

1 16 amino acids is noted in *S. obliquus* but this addition occurs in an external loop of the
2 protein and probably has no special function.

3 Until now, all Fe-hydrogenases possess a ferredoxin-like domain in the N-
4 terminus coordinating two [4Fe4S] clusters (FS4A, FS4B, as shown in Figure 3. Figure 3
5 is a schematic diagram showing the conserved cysteine residues and other important
6 amino acids of the H cluster. In Figure 3, the protein is illustrated as a *large gray arrow*.
7 *Small arrows* indicate parallelograms which demonstrate conserved amino acids in the
8 protein. Cysteines participating at the coordination of the [Fe-S] clusters are *gray*,
9 whereas identical amino acids are *black*. An insertion of 16 amino acids in the *S. obliquus*
10 protein is illustrated as a *spotted bar*. FS4 indicates the [4Fe-4S] cluster; and FS2
11 indicates the [2Fe-2S] cluster.

12 The iron sulfur cluster facilitates the transfer of electrons between external
13 electron donors or acceptors and the H-cluster. The N-terminus of the *S. obliquus* protein
14 is strongly reduced compared to other Fe-hydrogenases and no conserved cysteines are
15 found. Therefore it is postulated that all accessory Fe-S clusters (FS2, FS4A, FS4B,
16 FS4C) are missing. No indication of a second subunit has been observed during
17 purification of the protein.

18 In contrast to earlier observations in *S. obliquus*, the present inventor could
19 neither detect the postulated two subunits of a potential NiFe-hydrogenase, nor could he
20 find a Ni-dependency related to the hydrogenase activity. Francis reported about two
21 forms of hydrogenases in *S. obliquus* (*Photosynthetica*, **23**, 43-48 (1989)), but although

1 the present inventor used the same alga strain and identical adaptation conditions, a
2 second hydrogenase activity was not detected during the purification steps.

3 Physiological studies by others have shown that the hydrogen evolution is
4 coupled to the light reaction of the photosynthesis. In contrast to earlier observations in
5 *S. obliquus*, the measurement of PSII independent H²-production was not influenced by
6 DCMU. The electrons required for H²-evolution come from redox equivalents of the
7 fermentative metabolism and are supplied into the photosynthetic electron transport chain
8 via the plastoquinone pool.

9 For the first time, it is demonstrated that the ferredoxin PetF functions as the *in*
10 *vivo* electron donor of the Fe-hydrogenase from *S. obliquus*. Hydrogenase activity can be
11 specifically blocked by addition of the ferredoxin antagonist sulfo-DSPD (A. Trebst, *J*
12 *Methods Enzymol.*, **69**, 675-715 (1980)) and antibodies raised against the PetF protein. *In*
13 *vitro*, a hydrogen evolution by HydA was only measured with plant-type [2Fw-2S]
14 ferredoxins like PetF of *S. obliquus*, *C. reinhardtii* and spinach as electron mediators.
15 Bacterial Fe-hydrogenases are known to be reduced by [4Fe-4S] ferredoxins and do not
16 accept electrons from plant-type proteins (J. M. Moulis et al., *Biochemistry*, **34**, 16781-
17 16788 (1995)).

18 The analysis of the 3D-structure of the Fe-hydrogenase from *C. pasteurianum*
19 (CpI) gave evidence that the interaction with external electron donors might occur at the
20 accessory [Fe-S] clusters in the N-terminal domain (J. W. Peters, *Curr. Opin. Struct.*
21 *Biol.*, **9**, 670-676 (1999)). Based on the X-ray structure of CpI (N. Guex et al., *Trends*

1 *Biochem. Sci.*, **24**, 364-367 (1999), the Fe-hydrogenase of *S. obliquus* was modeled.
2 Figure 4 is a Schematic view of the structures of *S. obliquus* HydA (A), and the
3 electron donor ferredoxin (B). The figure shows the α carbons and the side chains of
4 charged residues that might be important for the electron transfer reaction or the
5 interaction between HydA and the ferredoxin from *S. vacuolatus*. The 16-amino acid
6 insertion of the hydrogenase appears as external loop and is distinguished as a *dotted line*.
7 The amino acid sequence of the mature HydA protein (His 19 –Tyr 404) was submitted
8 to the SWISS-MODEL server. The present inventor generated a model of HydA with the
9 known three-dimensional structure of the iron hydrogenase from *C. pasteurianum* as
10 template, sharing 57% sequence identity with the submitted sequence. The Protein Data
11 Bank file was visualized by the Swiss-PDB viewer (J. M. Moulis, *Biochemistry*, **34**,
12 16781-16788 (1995)). As shown in Fig. 4, a region of positive surface potential is
13 observed within HydA based on a local concentration of basic residues. In contrast to the
14 docking position of ferredoxin in Cpl, these charged amino acids in the *S. obliquus* Fe-
15 hydrogenase are located within the C-terminal domain, forming a niche for electron
16 donor fixation.

17 The known alga ferredoxin proteins exhibit high degrees of sequence identity
18 (over 85%) and the charged amino acids are strictly conserved. The petF sequence of *S.*
19 *obliquus* is unknown, but very recently the X-ray model of the ferredoxin from another
20 *Scenedesmus* species (*Scenedesmus vacuolatus*) was published (M. T. Bes et al.,
21 *Structure*, **7**, 1201-1213 (1999)). The structure revealed negatively charged amino acids

1 like aspartate and glutamate near the [2Fe-2S] cluster. The [Fe-S] center and the H-
2 cluster of the hydrogenase probably come into close proximity through electrostatic
3 interactions. This geometry is consistent with efficient electron transfer among these
4 prosthetic groups.

5 As already shown in various studies, a correlation exists between the duration of
6 time of the anaerobic adaptation and increase of hydrogen production. RT-PCR and
7 Northern blot analyses with mRNA of aerobic and anaerobically adapted cells from *S.*
8 *obliquus* showed an increased level of *hydA* transcript after one hour of induction.
9 Correspondingly, hydrogen evolution was only measured after a short time anaerobic
10 adaptation. These results suggest that the expression of the *hydA* gene is regulated at the
11 transcriptional level. The small amount of transcript that was detected at t=0 may be due
12 to transcript analysis induced by micro-anaerobic conditions during the RNA isolation
13 procedure. Alternatively, a low level of *hydA* transcript might be constitutively present in
14 the cell and is only drastically increased after anaerobic adaptation.

15 The foregoing discloses a monomeric enzyme, iron hydrogenase (HydA), having
16 a molecular mass of 44.5 kDa. (exclusive of the transit peptide associated therewith)
17 derived from *Scenedesmus obliquus*. The polypeptide derived from the cDNA sequence,
18 set forth herein as SEQ. ID. NO. 4, has a length of 448 amino acids and is the smallest
19 hydrogenase described to date. The nucleic acid sequence coding HydA in *Scenedesmus*
20 *obliquus* is set forth in SEQ. ID. NO. 1 appended hereto, and the cDNA sequence is set
21 forth as SEQ. ID. NO. 7.

1 In addition to the unicellular green algae *Scenedesmus obliquus*, discussed above,
2 other algae within the order of Chlorophyta (e.g. *Chlamydomonas reinhardtii* and
3 *Chlorella fusca*) contain a gene (*hydA*) coding for a novel iron-hydrogenase enzyme
4 (HydA) as will be discussed below. This gene, through its encoded enzyme, catalyzes the
5 synthesis of molecular hydrogen from protons and high potential energy electrons, and
6 releases significant amounts of hydrogen gas, which is a valuable and clean source of
7 energy. As with *Scenedesmus obliquus*, the process of H₂- production entails the
8 utilization of sunlight and the oxidation of water or organic substrate in photosynthesis to
9 generate reduced ferredoxin, which is the carrier of the high potential energy electrons.
10 The isolation, sequencing and characterization of the *hydA* genomic DNA, cDNA,
11 precursor and mature iron-hydrogenase of two additional photosynthetic eukaryotes
12 which may be used for hydrogen gas production is disclosed below.

13 **II. *Chlamydomonas reinhardtii***

14 ***C. reinhardtii* Algal Strains and Growth Conditions**

15 Wild-type *C. reinhardtii* 137c(mt+) strain was originally obtained from the
16 *Chlamydomonas* Culture Collection at Duke University. The strain was cultured
17 photoheterotrophically in batch cultures at 25 °C under continuous irradiance of 150
18 μmol photons per square meter per second. Cultures containing TAP (Tris acetate
19 phosphate) medium were flushed vigorously with air supplemented with 5% CO₂. Cells
20 were collected by centrifugation (8 minutes @ 5000g) in the mid-exponential growth
21 stage (1-2 x 10⁶ cells per ml). After harvesting the cells in the mid-exponential stage of

1 growth, the pellet was resuspended in 0.02 vol. of fresh TAP medium. The algae were
2 anaerobically adapted by flushing the culture with argon in the dark.

3 **Hydrogen Evolution Assay**

4 Hydrogenase activity of *C.reinhardtii* was determined in vitro with reduced
5 methyl viologen using a gas chromatograph (Hewlett Packard 5890 A Series II, column:
6 molecular Sieve 5 Å, Mesh 60/80). The assay, containing in a final volume of 2 mL Pipes
7 pH 6.8 (20 mM), Na₂S₂O₄ (20 mM), MV (5 mM), was incubated anaerobically at 25 °C
8 for 20 min. One unit is defined as the amount of hydrogenase evolving 1 µmol H₂ per
9 minute.

10 **Purification of the Fe-hydrogenase and Amino Acid Sequence**

11 Cells from a 40-L culture of *C. reinhardtii* were harvested by ultra filtration
12 through an Amicon Ultrafiltration System DC 10 LA, equipped with a hollow-fiber filter.
13 The pellet was resuspended in 200 mL TAP medium. After anaerobic adaptation by
14 flushing the solution with argon for 1 h in the dark, all steps were performed under
15 strictly anaerobic conditions. The isolated Fe-hydrogenase was chemically cleaved by
16 cyanogen bromide (CNBr). After separation of the CNBr fragments on an SDS
17 polyacrylamide gel, four peptides were blotted onto a poly(vinylidene difluoride)
18 membrane and were sequenced. Automated Edman degradation was performed with an
19 Applied Biosystem model 477 A sequencer with online analyser model 120 A.

20 **RNA Blot Hybridization**

1 Total nucleic acids were isolated from algae grown under aerobic conditions and
2 after anaerobic adaptation. Poly(A)+ RNA was isolated using the RNA Kit (Qiagen); 10
3 µg total RNA or 0.5 µg poly(A)+ RNA were separated on each lane of 1.2% agarose gels
4 in formaldehyde. The RNA was transferred to nylon membranes (Hybond+, Amersham)
5 and hybridized with RNA probes, which were labeled with digoxigenin (DIG)-dUTP by
6 *in vitro* transcription. Transcripts of the *hydA* gene were detected using a 1.0-kb *Sma* I
7 cDNA fragment. A DIG-dUTP labeled cDNA, which encodes the malate dehydrogenase,
8 was used as a control for a constitutive expressed gene. Fig. 5 (A) shows the structural
9 features of the *HydA* cDNA. Coding regions are marked as large arrows with the transit
10 peptide shown in black. Lines indicate 5' and 3' URT's. In Fig. 5 (B), the mosaic
11 structure of *HydA* is illustrated by gray (exons) and white (introns) boxes. The RNA and
12 DNA probes that were used for blotting are as noted on the figure.

13 **Suppression Subtractive Hybridization (SSH)**

14 SSH was performed with the Clontech PCR-select® cDNA Subtraction Kit
15 (Clontech Laboratories Inc., Palo Alto, CA, USA) according to the manufacturer's
16 recommendations, except for modifications of the PCR and hybridization conditions. The
17 mRNA was isolated from aerobically grown cells (driver) and from anaerobically
18 adapted algae (tester). The driver and tester cDNAs were denatured separately for the first
19 hybridization at 100 °C for 30 s and then incubated for 10 h at 68 °C. For the second
20 hybridization, driver cDNA was denatured at 100 °C for 30 s, then directly added to the
21 pooled mix of the previous hybridization, and incubated at 68 °C for 20 h. Primary and

1 secondary PCR conditions were altered to increase the specificity of the amplification.
 2 The PCR conditions with subtracted cDNA were as follows: 25 cycles each 94 °C for
 3 30 s, 68 °C for 30 s, and 72 °C for 1 min. The subtracted cDNA was subjected to a second
 4 round of nested PCR, using the same PCR conditions with a decreased number of 15
 5 cycles. Specific primers were used for the identification of the amplified *HydA* cDNA
 6 fragment. From the N-terminal amino-acid sequence, a degenerate oligonucleotide Hyd5
 7 [5'-GCCGCCCC(GC)GC(GCT)GC(GCT)GA(AG)GC-3'] was synthesized, taking into
 8 account known *C.reinhardtii* amino-acid sequences. The second primer, Hyd2 (5'-
 9 CCAACCAGGGCAGCAGCTGGTGAA-3'), was deduced from the conservative amino
 10 acid sequence motif of Fe-hydrogenases FTSCCP.

11 PCR was performed using either Hyd5 or Hyd2 and the nested PCR primer 2R
 12 from the Clontech Subtraction Kit. The PCR conditions were as follows: 20 pmol per ml
 13 of each primer were used; 35 cycles (denaturing at 95 °C for 40s, annealing at 54 °C for 1
 14 min, and extension at 72 °C for 1 min). The amplified cDNA fragments were cloned into
 15 the T overhang vector **pGEM® -T Easy** (Promega).

16 **Screening of the cDNA Library, Cloning and Sequencing**

17
 18 A cDNA library was constructed using the Stratagene ZAP Express cDNA
 19 synthesis Kit (Stratagene, La Jolla, CA, USA) with 5 µg mRNA of anaerobically adapted
 20 cells of *C. reinhardtii*. Double-stranded cDNA was ligated into the ZAP Express vector,
 21 packaged with the Gigapack Gold Kit, and transfected into *Escherichia coli* XL Blue

1 MRF cells. The primary recombinant library contained 5×10^6 recombinant phages and
2 was amplified according to the manufacturer's instructions.

3 A 366-bp PCR fragment was radiolabeled with [α - 32 P]dCTP using the random-
4 primer method. Approximately 5×10^5 plaques were analyzed under stringent
5 hybridization conditions, resulting in 20 positive signals. The pBK-CMV phagemid
6 vector with the different cDNAs was excised and used as a template for PCR, which was
7 performed by using Hyd2 and Hyd5 primers at an annealing temperature of 56 °C for 1
8 min. Four plasmids contained cDNA fragments that showed similarities to the 366-bp
9 fragment. All cDNA fragments were partially sequenced, and the largest clone, pAK60,
10 was completely sequenced. Sequencing was carried out by the dideoxy nucleotide
11 triphosphate chain-termination method using the T₇ sequencing Kit (Pharmacia Biotech).
12 Both strands of genomic and cDNA of hydA were completely sequenced using a nested
13 set of unidirectional deletions or hydA specific synthetic oligonucleotides. The sequences
14 of the Fe-hydrogenase are available under accession number CRE012098.

15 Primer extension experiments were performed as described previously by the
16 present inventor (*J. Biol. Chem.*, **276**, 6125-6132 (2001)) using a 22-mer oligonucleotide
17 (5'-AATAGGTGGTGCGATGAAGGAG-3'), which is complementary to the 5' end of
18 the hydA transcript.

19 **Expression studies in *E. coli* and Western blot analysis**

20 The coding region of hydA was amplified by PCR. The primers were identical to
21 the cDNA sequences coding for the N- and the C-terminus of the mature protein plus

1 several additional bases including *Nde*I and *Bam*HI restriction sites, respectively
2 (underlined). The oligonucleotide sequences were: HydNde (5'-
3 CATATGCCCGCACCCGCTGCGGAGGCGCCT-3'), HydBam (5'-CCGGATCC
4 TCAAGCCTCTGGCGCTCCTCA-3').

5 The *hydA* gene, corresponding to amino acids 57-497, was amplified, confirmed
6 by sequences analysis and cloned into corresponding sites of the pET9a expression vector
7 (Promega). The constructed plasmid was then transformed into *E. coli* strain BL21(DE3).
8 After induction with 1 mM isopropyl-thio- β -D-galactoside, the cells were resuspended
9 in lysis buffer. Crude extracts from *C. reinhardtii* were isolated by harvesting cells after
10 indicated anaerobic adaptation times. The pellet was resuspended in solubilization buffer
11 and incubated with vigorous vortexing at RT for 30 min. The protein extracts from *C.*
12 *reinhardtii* and *E. coli* were separated by 12%SDS/PAGE and blotted onto a
13 poly(vinylidene difluoride) membrane. Affinity-purified antibodies were diluted 1 :200
14 and used for Western blot analyses.

15 **Sequence analysis and protein modeling**

16 Nucleic acid and protein sequences were analyzed with the programs SCI ED
17 CENTRAL (Scientific Educational Software) and CLUSTALW. The BLAST server
18 (Altschul et al., *Nucleic Acid Res.*, **25**, 3389-3402 (1997)) of the National
19 Center for Biotechnology Information (Bethesda, MD, USA) was used for database
20 searches.

1 **Isolation of cDNA clones, which are differentially expressed during anaerobic**
2 **adaptation**

3 In order to amplify a part of the hydrogenase gene in a PCR reaction, degenerate
4 oligonucleotides corresponding to conserved regions of known Fe-hydrogenases were
5 used. All products of expected sizes were cloned and sequenced, but they showed no
6 homologies to other hydrogenases (data not shown). Examinations were then focused on
7 the process of anaerobic adaptation in *C. reinhardtii*, because the Fe-hydrogenase was
8 only detected under these conditions. The present inventor isolated two different
9 populations of mRNA and advantageously employed the SSH technique. Poly(A)+
10 RNA was isolated from aerobically grown *C. reinhardtii* cells and from a cell suspension
11 flushed 15 min with argon. After cDNA synthesis, subtractive hybridization, and PCR
12 experiments, the amplified PCR fragments were cloned and sequenced. Twenty different
13 clones containing inserts of 184-438 bp were analyzed. In transcription analyses, 15 of
14 them showed an increased signal under anaerobic conditions (data not shown). Database
15 comparisons (using GenBank/EBI DataBank) confirmed that eight of these cDNA
16 fragments are similar to genes encoding proteins of the cytoplasmic ribosome complex.
17 The sequences of six clones did not correspond to any entries in the databases.
18 Four of these novel clones showed differences in expression between aerobically grown
19 and anaerobically adapted cultures. Another cDNA fragment indicated similarity
20 to the 5' region of the Fe-hydrogenase from bacteria.

21 **Analysis of the *hydA* cDNA and genomic sequences**

1 A cDNA expression library was constructed using poly(A)+ RNA from
2 anaerobically adapted cells (15 min). Two oligonucleotides were generated on the basis of
3 the cDNA fragment isolated by SSH and the N-terminal sequences of the purified
4 hydrogenase. They were used to amplify a 366-bp cDNA fragment that showed 41%
5 identity to the corresponding part of the Fe-hydrogenase of *C. pasteurianum*. The
6 fragment was labeled with [α -³²P]dCTP and used to screen the cDNA library. Four
7 independent cDNA clones with different sizes of 2.4-, 1.9-, 1.7- and 1.6-kb were
8 identified and sequenced. The nucleotide sequence of the largest clone, 2399-bp, revealed
9 an ORF encoding a polypeptide of 497 amino acids. The cDNA also contained a 5'UTR
10 (158-bp) and a longer 3'UTR (747-bp excluding the polyadenylated tail). Characteristic
11 features of other *C. reinhardtii* cDNA clones, e.g. a high average G/C content (62.1%) and
12 a putative polyadenylation signal (TGTAAG) 727-bp downstream of the stop codon were
13 found. The transcription start position was confirmed by primer extension 158-bp
14 upstream of the ATG start codon.

15 Figure 5 (A) is a schematic map of the cDNA and the genomic DNA region of
16 HydA from *C. reinhardtii*, showing the structural features of the HydA cDNA. Coding
17 regions are marked as large arrows with the transit peptide shown in black. Lines indicate
18 5' and 3' UTRs. Fig. 5 (B) is a schematic map of the cDNA and the genomic DNA region
19 of HydA from *C. reinhardtii*, showing the structural features of the HydA cDNA. In Fig. 5
20 (B), the mosaic structure of hydA is illustrated by gray (exons) and white (introns) boxes.

1 The RNA and DNA probes that were used for the blotting experiments are noted on the
2 Figure.

3 Approximately 5-kb of the *hydA* genomic region was determined. The coding
4 sequence is interrupted by seven introns with sequences at their 5' and 3' ends,
5 corresponding to the typical splicing sequences from eukaryotes as shown in Figure 5B.
6 The promoter region does not contain a putative TATA box or any other known
7 transcription motifs. The sequence data were submitted to the GenBank/EBI DataBank
8 under accession number CRE012098. In subsequent studies, parts of the cDNA sequence
9 were determined by another group and deposited under accession number AF289201.

10 Southern hybridization experiments were performed at high stringency using a
11 PCR fragment as probe. They showed the presence of one hybridizing signal of
12 similar intensity in different digestions, suggesting that *HydA* is encoded by a single copy
13 gene in the *C. reinhardtii* genome. The same hybridization pattern was observed even
14 under low stringency conditions (hybridization temperature 50 °C; data not shown).

15 **Characterization of the Fe-hydrogenase *HydA* from *C. reinhardtii***

16 The mature polypeptide consists of 441 amino acids with a calculated molecular
17 mass of 47.5 kDa and a predicted isoelectric point of 5.6. The N-terminal 56 amino acids
18 probably function as a transit peptide, because they show characteristics of polypeptides
19 that route proteins into the chloroplast stroma. The stromal targeting domain is
20 most likely cleaved by a stromal peptidase at the conserved cleavage motive Val-Ala-
21 Cys-Ala. In addition to the detection of the protein using antibodies raised against the

Fe-hydrogenase, the localization of the mature protein in the chloroplast stroma is indicated by a high content of hydroxylated and basic amino acids in the transit peptide sequence.

The deduced amino-acid sequence of the mature HydA polypeptide from *C. reinhardtii* shows 60% identity and 71% similarity to the Fe-hydrogenase of *S. obliquus*. Comparisons with NiFe-hydrogenases of bacteria (including the photosynthetic cyanobacteria) had obviously lower scores, e.g. 25% similarity with the NiFe-hydrogenase (HoxH) of *Ralstonia eutropha*. A conserved domain of about 300 amino acids is found in the C-terminal part of all Fe-hydrogenases. The sequences are highly conserved, especially in the region that is involved in the catalytic mechanism (H-cluster), indicating structural similarity between Fe-hydrogenases. Four cysteine residues at positions 114, 169, 361 and 365 might coordinate the H-cluster in *C. reinhardtii*. Twelve strictly conserved amino acids of HydA proteins probably define a binding pocket surrounding the active center as shown by structural data reported by others for *C. pasteurianum* and *D. desulfuricans* iron hydrogenase. All of them are present in the *C. reinhardtii* protein (Pro37, Ala38, Thr74, Ala78, Cys113, Pro138, Met167, Lys172, Glu175, Phe234, Val240 and Met359). An interesting insertion of 45 amino acids was only identified at the C-terminus of the *C. reinhardtii* polypeptide (position 285-329).

The N-terminal region of the green algae protein is much shorter and completely different than all known Fe-hydrogenases. Amino-acid sequence analyses have indicated

that Fe-hydrogenases, in general, contain two [4Fe-4S] clusters (F-cluster) in a ferredoxin-like domain. They might be involved in the transfer of electrons from the donor to the catalytic center. This N-terminal domain with the F-cluster or other conserved cysteines is completely missing in HydA of *C. reinhardtii*. A novel electron transport pathway is postulated from the exogenous donor (ferredoxin) directly to the H-cluster.

Protein sequencing of the enzyme and recombinant expression of HydA in *E. coli*

To verify that the *hydA* ORF encodes the Fe-hydrogenase of *C. reinhardtii*, the enzyme was purified according to Happe and Naber (*Eur. J. Biochem.* **214**, 475-481 (1993)). The purified protein was able to evolve hydrogen, when incubated with reduced methyl viologen. After proteolytic digestion with cyanogen bromide, four bands of 4, 8, 9 and 11 kDa were detected after SDS/PAGE separation (data not shown). Two fragments (9 and 11 kDa) were sequenced by Edman degradation. Figure 6 shows the nucleotide sequence of the *hydA* cDNA and the deduced amino acid sequence of the hydrogenase from *C. reinhardtii*. The two fragment sequences are identical with the deduced amino acid sequence of *hydA* (sequences are shadowed in gray in Fig.6). The fragment corresponding to the cDNA region between 158 and 1636 bp of *hydA* was *NdeI*-*Bam*HI cloned into the expression vector pET9a. The heterologously expressed protein was detected using antibodies raised against the Fe-hydrogenase. Both the purified Fe-hydrogenase of *C. reinhardtii* and the overexpressed enzyme had the same size

(47.5 kDa). No hydrogenase activity could be measured within the lysate of the induced *E. coli* cells. This result is in agreement with Stokkermans et al. and Voordouw et al. (*ibid*) who also detected no H₂ -production of the recombinant expressed Fe-hydrogenase from *Desulfovibrio vulgaris* in *E. coli* cells. An explanation might be the inability of *E. coli* to assemble the unique active site of the Fe-hydrogenases. It is known that *E. coli* has only three NiFe-hydrogenases with a different maturation system for the catalytic center.

Figure 6 shows the nucleotide sequence of the *hydA* cDNA and the deduced amino acid sequence of the hydrogenase from *C. reinhardtii*. The sequence was submitted to the GenBank/EBI Data Bank under accession number CRE012098. A curved arrow marks the transcription start point. The ATG start codon and the TGA stop codon are drawn in boxes. Boldface letters indicate the cDNA sequence. Gray shadows mark amino acids corresponding to polypeptide sequences that were determined by sequencing the N-terminus of the protein. Black shadows mark the putative transit peptide, and the underlined amino acids indicate the putative cleavage site for the endopeptidase. Boldface double underlined letters indicate a signal for polyadenylation.

Induction of gene expression during anaerobic adaptation

In aerobically grown cells, neither hydrogenase activity nor protein can be identified by immunoblot analysis. However, HydA can be detected only 15 min after anaerobic adaptation. The expression of the *hydA* gene is probably regulated at

1 the transcriptional level. Total RNA was isolated from cells that had been anaerobically
2 adapted by flushing with argon for 0,15, and 30 min. Northern blot hybridization
3 demonstrated that the hydA gene is expressed very rapidly after the beginning of
4 anaerobic adaptation. No transcript could be detected before adaptation ($t = 0$), but a
5 significant signal occurred after just 15 min of anaerobiosis. The size of the transcripts
6 (2.4 kb) confirmed the full-length of the isolated hydA cDNA fragment.

7 **Differentially expressed genes during anaerobic adaptation**

8 In the light, algae degrade cellular starch via glycolysis and hydrogen gas is
9 evolved. It has been suggested that reducing equivalents from the glycolysis
10 or the citric acid cycle can transfer their electrons to the photosynthetic electron transport
11 chain (M. Gibbs et al., *Plant Physiol.* **82**, 160-166 (1986)). However, the molecular
12 principles of the gene induction under anaerobic conditions in *C. reinhardtii* are poorly
13 understood.

14 The present inventor has investigated the patterns of gene expression in
15 aerobically grown and anaerobically adapted cells by isolating differentially expressed
16 genes. The SSH method combines subtractive hybridization with PCR to generate a
17 population of PCR fragments enriched with gene sequences that are only expressed under
18 anaerobic conditions. Compared to other PCR-based cloning strategies, such as
19 differential display, the great advantage of SSH is that fewer false positives are generated;
20 70% of the cloned fragments represented differentially expressed genes.

1 Among the 20 sequenced cDNA clones, three DNA fragments encoding the
2 ribosomal S8 protein were found. Most of the other sequences (eight of 20) also
3 corresponded to ribosomal protein sequences. This might indicate that the transcripts of
4 the ribosomal protein genes (rps, rpl) accumulate under stress conditions. This is in good
5 agreement with Dumont et al. (*Plant Sci.*, **89**, 55-67 (1993)) who found that an
6 accumulation of ribosomal proteins takes place under phosphate starvation. Moreover,
7 two of the identified cDNAs encode for proteins, (aldolase, enolase), which are induced
8 in other organisms by anaerobic stress. Anaerobic treatment of maize seedlings alters the
9 profile of total protein synthesis. It is known that the induction of the anaerobic proteins
10 is the result of an increased mRNA level. Maize (*Zea mays* L.) responds to anaerobic
11 stress by redirecting the synthetic machinery towards the synthesis of some enzymes
12 involved in glycolysis or sugar-phosphate metabolism.

13 ***C. reinhardtii* HydA belongs to a new class of Fe-hydrogenases**

14 HydA of *C. reinhardtii*, the first isolated gene encoding a hydrogenase of a
15 photosynthetic eukaryotic cell, represents a novel type of Fe-hydrogenases. Parts of the
16 deduced amino-acid sequence of the cDNA correspond to the polypeptide sequence of the
17 tryptic fragment (VPAPGSKFEELLKHRAAARA), and the N-terminus
18 (AAPAAAEAPLSHVQQALAEAKPKD) from the purified native enzyme. Further
19 evidence that the isolated cDNA encodes an Fe-hydrogenase is the fact that the
20 recombinant HydA specifically reacts with the antibodies raised against the active
21 enzyme. The amino-acid sequence of HydA shows only considerable similarity to Fe-

1 hydrogenases but not to NiFe-hydrogenases. The Fe-hydrogenase family is one class of
 2 hydrogenases defined by Vignais et al. (*FEMS Microbiol. Rev.* **25**, 455-501 (2001)). The
 3 enzymes have been identified in a small group of anaerobic microbes, where they often
 4 catalyze the reduction of protons with a high specific activity to yield hydrogen.
 5 Interestingly, Fe-hydrogenases were not found in cyanobacteria, the free-living ancestor
 6 of plastids, suggesting a noncyanobacterial origin for the algal hydrogenases.

7 The important structural features found among the amino-acid sequences of Fe-
 8 hydrogenases are also present in the *C. reinhardtii* hydrogenase sequence. A highly
 9 conserved domain of about 130 amino acids was detected in the C-terminal part of the
 10 protein. The designated active site domain consists of an atypical [Fe-S] cluster
 11 (H-cluster). In *C. pasteurianum*, the H-cluster contains six Fe atoms arranged as a [4Fe-
 12 4S] subcluster bridged to a [2Fe] subcluster by a single cysteinyl sulfur. The [4Fe-4S]
 13 subcluster is coordinated to the protein by four cysteine ligands, which have also been
 14 found in the amino acid sequence of *C. reinhardtii*. A number of mostly hydrophobic
 15 amino acid residues define the environment of the active site and might have a function in
 16 protecting the H-cluster from solvent access. In contrast to all Fe-hydrogenases, including
 17 HydA of *S. obliquus*, the enzyme of *C. reinhardtii* has an interesting additional
 18 protein domain. A small insertion of 45 amino acids between residue Ser284 and Val330
 19 builds an external loop of the protein that might be involved in electrostatic binding
 20 of the natural electron donor ferredoxin (Happe et al., unpublished results) .

In the N-terminus of other Fe-hydrogenases, further cysteine residues were found that bind accessory iron sulfur clusters. Others have shown that a ferredoxin homologous domain (F-cluster) coordinates two [4Fe-4S] clusters in all non-algal Fe-hydrogenases. An additional [4Fe-4S] cluster and one [2Fe-2S] center were detected within the Fe-hydrogenases of *C. pasteurianum* (Peters et al., *Science*, **282**, 1853-1858 (1998)). Based on similarities of the primary sequences, the same cofactors are proposed for *Thermotoga maritime* (A. Akhmanova et al., *Nature*, **396**, 527-528 (1998)) and *Nyctotherus ovalis* (J. Meyer et al., *Biochim. Biophys. Acta*, **1412**, 212-229 (1999)). The F-cluster is responsible for the electron transfer from the electron donor (mostly ferredoxin) to the active center. It has been suggested by Vignais et al. (*FEMS Microbiol. Rev.*, **25**, 455-501 (2001)) that the proteins containing two F-clusters are ancestors of the Fe-hydrogenases.

The N-terminus of the *C. reinhardtii* and *S. obliquus* proteins is strongly reduced, and conserved cysteines were also not found. Therefore it is suggested that all accessory [Fe-S] clusters are missing in the algal hydrogenases. The native protein of *C. reinhardtii* is located in the chloroplast stroma. The first 56 amino acids of the unprocessed enzyme probably function as a transit peptide, because they were not characterized in the purified hydrogenase and a putative peptidase cleavage site (Val-Ala-Cys-Ala) could be detected at the end of this fragment.

The natural electron donor of the hydrogenase in *C. reinhardtii* is the ferredoxin (PetF) of the photosynthetic electron transport pathway. Measuring the H₂-evolution, it has been found that the hydrogenase activity is directly linked to the 47.5 kDa subunit. As

1 a second subunit necessary for hydrogenase activity has not been found, it is suggested
2 that a direct electron transfer from PetF to HydA takes place. *In vitro*, a hydrogen
3 evolution by HydA was only measured with plant-type [2Fe-2S] ferredoxins such as PetF
4 of *C. reinhardtii*, *S. obliquus* and spinach as electron mediators (data not shown).

5 Figure 7 is a schematic diagram illustrating the light-dependent photoevolution of
6 hydrogen in green algae. The electrons for hydrogen evolution are fed into the
7 photosynthetic electron transport chain either via PS II or via the plastoquinone pool after
8 oxidation of reducing equivalents. The natural electron donor, PetF, transfers the
9 electrons from PS I to the hydrogenase.

10 The most likely explanation for photosynthetic green algae retaining the
11 anaerobically induced hydrogenases is that the enzymes ensure the survival of the cells
12 under anaerobic conditions. Melis et al. have shown that H₂-evolution is the only
13 mechanism available to the algae for generating sufficient amounts of ATP under
14 S-depleted anaerobic conditions. It is known that *C. reinhardtii* is still able to
15 photoproduce hydrogen when photosystem II is inhibited by DCMU, but no H₂-evolu-
16 tion occurs after an addition of 2,5-dibromo-3-methyl-6-isopropyl-p -benzoquinon
17 (DBMIB; Figure 7). Under anaerobic conditions, accumulated reducing equivalents from
18 the fermentative metabolism cannot be oxidized via respiration, as the electron acceptor
19 oxygen is missing. The NAD(P)H reductase protein complex has recently been
20 isolated from plants, and inhibitor experiments have shown evidence of a membrane-
21 bound, chloroplast-located reductase in *C. reinhardtii* (D. Godde et al., *Arch. Microbiol.*

1 127, 245-252 (1980)). The light-dependent electron transport of the H₂-evolution is
2 driven by plastoquinone and photosystem I. The donor ferredoxin transfers electrons
3 to the hydrogenase in a final step and molecular hydrogen is released (Figure 7).

4 **Regulation of *hydA* at the transcriptional level**

5 Studies indicate that there is a correlation between the increase of hydrogen
6 production and the anaerobic adaptation, which was documented by activity
7 measurements (T. Happe et al., *Eur. J. Biochem.* **222**, 769-775 (1994)), and immunoblots.
8 It is likely that the induction of *hydA* is regulated on the level of transcription. It is
9 observed that the amount of mRNA increased directly with the measured H₂-evolution. In
10 *C. reinhardtii*, a dramatic change in the hydrogenase transcript level occurs during the
11 shift from an aerobic to an anaerobic atmosphere, which means that the transcription is
12 regulated by the oxygen status of the cells. A very rapid increase of the *hydA* transcript
13 was detected in the first 30 min of anaerobiosis. This quick increase of gene transcription
14 is only reported for the *cyc6* gene in *C. reinhardtii* (K.L. Hill et al., *J. Biol. Chem.* **266**,
15 15060-15067 (1991)) and for the SAUR (Small Auxin-Up RNA) genes in plants (Y. Li et
16 al., *Plant Physiol.* **106**, 37-43 (1994)). Interestingly, the *hydA* gene of *S. obliquus* is
17 constitutively transcribed under aerobic conditions indicating another regulation system
18 for the expression of the hydrogenase. At the moment, it is not clear if his effect rests
19 upon a new synthesis or a higher stability of the *hydA* mRNA.

20 As with other nuclear genes, the promoter region of the *hydA* from *C. reinhardtii*
21 contains no conserved TATA box or other motif similarities. As no defined motif

structures in the promoter region of *hydA* have been found, further genetic analyses are necessary to investigate the rapid induction of *hydA* in *C. reinhardtii*.

III. *Chlorella fusca*

The isolation and molecular characterization of the Fe-hydrogenase from the unicellular green alga *Chlorella fusca* was also performed. Hydrogenase activity was observed in a culture of the unicellular green alga *Chlorella fusca* after anaerobic incubation, but not in the related species *Chlorella vulgaris*. Specific PCR techniques lead to the isolation of the cDNA and the genomic DNA of a special type of [Fe]-hydrogenase in *C. fusca*. The functional Fe-hydrogenase was purified to homogeneity as described above, and its N-terminus was sequenced. The polypeptide sequence shows a high degree of identity with the amino acid sequence deduced from the respective cDNA region. Structural and biochemical analyses indicate that ferredoxin is the main physiological electron donor to this [Fe]-hydrogenase. The nucleotide sequence reported herein has been submitted to the GenBank/EBI Data Bank with accession number AJ298227. The nucleic acid sequence of the genomic DNA (3,290 bp) of *C. fusca* is set forth in SEQ. ID. NO. 3, and the amino acid sequence of the precursor protein (436 amino acids) is set forth in SEQ. ID. NO. 6. The cDNA sequence is presented in SEQ. ID. NO. 9.

The transcription of the iron-hydrogenase is very rapidly induced during anaerobic adaptation of the green algae. Hydrogen photoproduction by the cells can be observed soon following this induction. The genomic, cDNA and polypeptide sequences

of three representative green algae are offered as examples of the properties of the *HydA* gene and of the enzyme that it encodes. These genomic, cDNA and polypeptide sequences from three representative green algae are also offered as examples of the potential application of the *hydA* gene from Chlorophyta in the process of commercial hydrogen production:

Scenedesmus obliquus

Genomic DNA: 5,001 bp, SEQ. ID NO. 1 (attached hereto)
 cDNA: 2,636 bp SEQ. ID NO. 7 (attached)
 Precursor protein: 448 amino acids, 44.5 kD, SEQ. ID NO. 4 (attached hereto)

Chlamydomonas reinhardtii

Genomic DNA: 5,208 bp, SEQ. ID. NO. 2 (attached)
 cDNA: 2,399 bp, SEQ. ID. NO. 8 (attached)
 Precursor protein: 497 amino acids, 53.1 kD, SEQ. ID. NO. 5 (attached)

Chlorella fusca

Genomic DNA: 3,265 bp, SEQ. ID. NO. 3 (attached)
 cDNA: 2,421 bp, SEQ. ID NO. 9 (attached)
 Precursor protein: 436 amino acids, SEQ. ID. NO. 6 (attached)

This new class of iron-hydrogenases has a C-terminal portion and active site region (H-cluster) similar to that reported in non-photosynthetic prokaryotes (e.g. *Clostridium pasteurianum*). Cysteine residues and distinct other amino acids which are strictly conserved in the active site (H-cluster) of such non-photosynthetic prokaryotes

1 are also conserved in the iron-hydrogenase of green algae. However, the N-terminal
2 region of the green alga iron-hydrogenase is substantially different from that of HydA in
3 non-photosynthetic prokaryotes, revealing novel and unobvious pathways of electron
4 transport for photosynthetic hydrogen production in green algae.

5 Distinct iron-sulfur [Fe-S] centers, referred to as FS2, FS4C, FS4B and FS4A, are
6 encountered in the N-terminal region of the iron-hydrogenase in non-photosynthetic
7 prokaryotes, and are thought to be instrumental in the transport of high potential energy
8 electrons from bacterial ferredoxins to the catalytic site of the H-cluster of the
9 hydrogenase. These distinct FS2, FS4C, FS4B and FS4A [Fe-S] centers are missing from
10 the HydA of green algae. A mature-protein folding-model of the green alga iron-
11 hydrogenase and analysis of its structure revealed a protein region of positive surface
12 potential, evidenced by the presence of basic amino acids, which are uniquely localized
13 within the C-terminal domain and, therefore, near the catalytic site of the H-cluster. On
14 the other hand, a model of the structure of green alga ferredoxin, which is different from
15 prokaryotic ferredoxins, revealed negatively charged amino acids near the [2Fe-2S]
16 electron donor site of this molecule. Structural analysis revealed that the [2Fe-2S] center
17 of green algal ferredoxin and the H-cluster of the hydrogenase probably come into close
18 proximity through electrostatic interactions. This molecular geometry is consistent with a
19 direct and efficient electron transfer between these two prosthetic groups. Thus, the *hydA*
20 gene of green algae encodes an iron-hydrogenase polypeptide with a novel structure, one
21 that uniquely permits a direct coupling and efficient electron-transfer from a [2Fe-2S]

